Bordetella pertussis evolution in the (functional) genomics era.

Thomas Belcher and Andrew Preston*.

Department of Biology and Biochemistry, University of Bath, Bath, U.K.
Corresponding author. Dr Andrew Preston, Department of Biology and Biochemistry, University of Bath, Bath, BA2 7AY, U.K. a.preston@bath.ac.uk.

Running title.
B. pertussis evolution.

Abstract.
The incidence of whooping cough caused by *Bordetella pertussis* in many developed countries has risen dramatically in recent years. This has been linked to the use of an acellular pertussis vaccine. In addition, it is thought that *B. pertussis* is adapting under acellular vaccine mediated immune selection pressure, towards vaccine escape. Genomics-based approaches have revolutionised the ability to resolve the fine structure of the global *B. pertussis* population and its evolution during the era of vaccination. Here, we discuss the current picture of *B. pertussis* evolution and diversity in the light of the current resurgence, highlight import questions raised by recent studies in this area and discuss the role that functional genomics can play in addressing current knowledge gaps.

Resurgence of Pertussis.

Whooping cough, or Pertussis, caused by infection by *Bordetella pertussis,* was once a major cause of infant deaths worldwide. The introduction of Pertussis vaccination during the 1940's and 1950's in developed countries appeared to largely eradicate the disease in these countries. Pertussis remained a prominent infection in developing countries but since 1980, the percentage of the target population receiving at least one vaccination has risen from 30 to 90%, and the number of reported cases has fallen from nearly 2 million to approximately 161 500 per year in that time (WHO 2014). However, while a vast majority of developed countries switched from the use of whole cell Pertussis vaccines (WCV) to acellular (ACV) ones, developing countries continue to use WCV-based combination vaccines.

It is thus of great concern that Pertussis appears resurgent in numerous countries despite high levels of vaccination. There is a strong link between the use of ACVs and resurgence, which has raised a number of important questions regarding vaccine efficacy and the ability of these vaccines to continue to control disease at the population level (Burns, et al. 2014: S32-S5). This is not only of concern to the developed countries using these vaccines, but world-wide as many other countries aspire to switch to ACVs due to their perceived better safety profile compared to WCV.

ACVs comprise purified *B. pertussis* proteins, most contain filamentous haemagglutinin, pertussis toxin and pertactin while some also include fimbrial proteins (Fim2 and Fim3). ACVs were developed in response to concerns over reactions to vaccination with WCV. Injection site soreness, irritability and low grade fever were common sequelae while a syndrome termed Pertussis-vaccine encephalopathy was an association between WCV and serious neurological episodes (Cherry 1992: 1033-8, Cherry 1996: S 259-S 63). Although, now largely disproved, at the time this association was extremely damaging to public confidence in Pertussis vaccination. Compared to WCV, ACVs are considered much better tolerated (Cherry 1993: 21-4).

However, it has been long-recognised that WCV and ACV induce immunity of different profiles even though clinical trials of ACVs established that they are as efficacious as WCV for the prime consideration of protecting young babies and infants from classic Pertussis disease, for example (Lugauer, et al. 2002: 142-6). However, it is now becoming evident that there are important differences in the duration of immunity induced by the two vaccines, with ACV-induced immunity waning earlier than that induced by WCV (Rendi-Wagner, et al. 2006: 5960-5, Witt, et al. 2012: 1730-5). These studies found a strong correlation between the use of ACV and a higher risk of infection in older children and adolescents.

Also there are clear differences between the T cell responses induced by WCV compared to ACV, being Th1/Th17-biased compared to Th2/Th17 biased respectively (Ross, et al. 2013: e1003264). The consequences of differences between ACV and WCV induced immunity have become clearer in a recent study in the recently developed infant baboon model of infection. This demonstrated that ACVs protect the individual from manifestations of serious disease but not from colonisation and that ACV-vaccinated, colonised baboons are able to transmit bacteria to naive animals (Warfel, et al. 2013). The strongest resistance to challenge with *B. pertussis* was observed for convalescent animals while baboons vaccinated with WCV became colonised but to a lower extent, and for a shorter period of time, than ACV vaccinated animals. The same skew in T cell responses to ACV were observed in baboons.

Overall, there are clear differences in the type of immunity that ACVs induce compared to WCV, involving both B and T cell mediated immunity. The results of this are a shorter duration of immunity and a decrease in protection from infection with ACVs compared to WCV. Readers are referred to an excellent recent review for a detailed discussion of this (Warfel and Edwards 2015: 48-54). These data strongly suggest that the switch from WCV to ACV will have caused changes to the epidemiology of Pertussis in regions using ACV. Thus, it is likely that this has at least played a part in, if not caused, the resurgence of Pertussis experienced by many of these countries.

It is plausible that a change in host immunity and overall epidemiology will have an effect on the *B. pertussis* population. It is argued that *B. pertussis* is evolving under the different immune-selection pressure induced by ACVs (Mooi 2010: 36-49, Mooi, et al. 2014: 685-94) and that pathogen adaptation towards escape from vaccine-mediated immunity is also part of the resurgence story. To this end, a number of studies have identified subtle genetic changes in *B. pertussis* isolates over time, consistent with this hypothesis (see below). However, definitive evidence for current *B. pertussis* strains having altered infection biology is lacking. In fact the current debate regarding possible causes of, and solutions to Pertussis resurgence has highlighted that understanding of *B. pertussis* infection biology has many critical knowledge gaps that will need to be filled to meet these aims.

Much of our knowledge of *B. pertussis* virulence is based on a relatively small number of *B. pertussis* proteins. The advance of functional genomics means it is now feasible to study *B. pertussis* at a systems level, using genome-wide approaches that are unbiased compared to more traditional studies of individual proteins or genetic loci. Here we discuss the contribution of genomics analyses to the current picture of *B. pertussis* adaptation, and how they might address the important questions arising during debate of resurgence.

B. pertussis **phylogeny in the pre-genomics era.**

The study of *B. pertussis* phylogeny in the pre-genomics era was hampered by the very low levels of variation between strains. *B. pertussis* lacks phenotypic characters with the variability to be used for differentiation, such as the surface antigens used for serotyping of many of other bacterial pathogens.

PFGE. Pulse field gel electrophoresis has been used widely to differentiate between *B. pertussis* strains. It benefits from not requiring knowledge of particular bacterial components but suffers from difficulty in ensuring reproducibility between different labs and limited discriminatory power. However, from its use it was evident that *B. pertussis* isolates were divisible into different PFGE types, and that there was relatively limited diversity in the number of types observed given the seemingly large scope for variation, for example see (van Gent, et al. 2015: 821-30). It is not clear how many types arose by mutation of the restriction enzyme recognition sites or through intra-genome recombination and thus PFGE is unable to reveal information about the nature of differences between strains with different PFGE types.

MVLA. Multiple-locus variable number tandem repeat analysis (MVLA) discriminates between strains based on variation in the number of repeats at distinct loci. These regions are amplified by PCR and the size of the amplicon used to determine the number of repeats at each locus within strains. Strains are assigned to a MLVA type according to their profile of repeat sizes. This approach has identified numerous different MLVA types, for example see (Kurniawan, et al. 2010:

297-300). It is high throughput allowing analysis of a large number of strains but suffers from it not being clear as to how, or how often, variation in repeats occur. Thus, a clone in which a particular repeat length changes will be assigned as a different MVLA type even though the remainder of the genome may be identical. Thus, it is not always apparent as to what the MLVA types of a collection of strains indicate about their relatedness.

Allele typing. Many studies described strains as allele types, based on identification among strains of a limited number of alleles of the genes encoding a small number of key virulence factors, notably pertactin (*prn*), *ptxA* (encoding the enzymatically active subunit of pertussis toxin) and fimbrial subunits (*fim*) type 2 and 3. This was later expanded to include alleles of the promoter region of the pertussis toxin locus (*ptxP*). The profile of alleles of strains was used to assign them to a Sequence Type, for example see (Litt, et al. 2009: 680-8). However, the limited variation among *B. pertussis* meant that focusing on such a small portion of the genome represented by these few genes resulted in many strains appearing identical, with changes in Sequence Types being identified relatively infrequently.

The *ptxP3* allele has received particular attention. The SNP giving rise to this allele lies in a binding site for the transcriptional regulator BvgA and is thought to increase the affinity of BvgA for binding to the promoter and in doing so, increase the level of transcription of the associated pertussis toxin locus (Mooi, et al. 2009: 1206-13). Thus, strains carrying the *ptxP3* allele secrete increased amounts of toxin compared to strains carrying other alleles and this has been proposed to increase transmission of *ptxP3* strains, possibly through increasing the severity of clinical symptoms (Mooi, et al. 2009: 1206-13). Certainly the frequency of *ptxP3* has increased dramatically since the mid 1990's such that a majority of current isolates in developed countries carry *ptxP3* (Lam, et al. 2012: 492-5). A recent study identified that both the SNP itself, and the genetic background of a *ptxP3* strain contributed to an increase in fitness in a mouse model of infection compared to a *ptxP1* strain (King, et al. 2013: e66150) supporting the hypothesis that *ptxP3* strains represent a lineage with a fitness advantage over other strains, although the exact mechanism operating is not fully understood.

Such studies revealed that new alleles or sequence types arose over time and often one allele appeared to replace previously dominant alleles across the *B. pertussis* population, leading to the picture in which a particular clone appeared to dominate at any one time until replaced by a novel type, termed selective sweeps (van Gent, et al. 2012: e46407). These approaches were very valuable for tracing the replacement of clones and led to the realisation that the *B. pertussis* population changed over time. There is evidence from the case of the *ptxP3* allele that some of these changes may reflect changes in fitness of *B. pertussis* and as certain clonal sweeps coincided with changes to vaccine schedules or the switch from WCV to ACV, this raised the possibility that vaccination was a driver of *B. pertussis* evolution.

However, the focus on very limited regions of the genome, in a bacterium with relatively low levels of mutation, meant that there was insufficient ability to discriminate between strains co-circulating in most populations to reveal the true picture of strain diversity of *B. pertussis*.

Genomic analysis.

Genomics has been hugely informative to the study of *Bordetella, B. pertussis* in particular. The original *Bordetella* genome project simultaneously sequenced the genomes of a representative strain of *B. pertussis*, *B. parapertussis* and *B. bronchiseptica* and revealed much about the speciation of *B. pertussis* and *B. parapertussis* from *B. bronchiseptica* (or a *B. bronchiseptica*-like ancestor) (Parkhill, et al. 2003: 32-40), identifying that *B. pertussis* evolved as a species primarily by loss of DNA and intra-genomic recombination arising via recombination between IS element repeats, particularly IS481 (Parkhill, et al. 2003: 32-40). A notable feature of *B. pertussis* is the very high number of copies of these repeats, particularly IS481 that is present at around 250 copies per genome. IS1663 is present at around 30 copies. Such deletions have resulted in the genomes of *B. pertussis* being over a Mb smaller than those of *B. bronchiseptica*. On going recombination between IS element repeats is still shaping the *B. pertussis* genome (see below).

The monomorphic nature of *B. pertussis* has been taken as evidence that it arose very recently although its origin has been debated. There is no evidence in the ancient literature for the presence of classic whooping cough. Such references begin to appear in medieval Europe suggesting that classic whooping cough became established in Europe around the 14th or 15th century (Lapin 1943), possibly via import from other regions. The expansion of IS elements in *B. pertussis* is suggestive of a niche jump, as the fixation in the population of what would be ordinarily a highly deleterious level of genome disruption is indicative of a lack of competition from progenitor bacteria, typical of break through into a new niche (Preston, et al. 2004: 379-90). Wider analysis of the expansion of IS elements in niche restricted bacteria point towards the specialisation of human-specific bacterial pathogens occurring during the Neolithic period (approximately 10 000 years ago) due to the advent of agriculture creating stable and larger human communities that are favourable environments for bacteria to specialise (Mira, et al. 2006: 200-6).

While this niche jump is suggestive of a recent introduction of the Bordetellae into humans, analysis of *B. bronchiseptica* identified a distinct clade (Cluster IV) of seemingly human-associated *B. bronchiseptica* from which *B. pertussis* arose (Diavatopoulos, et al. 2005: e45). Thus, the genomic changes in *B. pertussis* likely represent the specialisation to a human-restricted niche, rather than the introduction into humans per se. It is unclear if this was concomitant with the evolution of the virulence represented by classic whooping cough, or if this was a later event arising from other genomic changes. For example, pertussis toxin appears important for virulence due to its subversive effect on immune cells. The genetic locus for PT and its secretion is present also in *B. bronchiseptica* and *B. parapertussis*, suggesting that it was present in the ancestor of *B. pertussis*. However, toxin expression is observed only in *B. pertussis*. While two mutations in toxin genes in *B. parapertussis* may explain its lack of toxin production, the locus appears intact in *B. bronchiseptica*. The promoter for this locus is identical between *B. bronchiseptica* and *B. parapertussis*. While a number of *ptxP* alleles have been identified among *B. pertussis*, including *ptxp3* (and see below), all contain several differences from the *B. bronchiseptica* promoter that likely enhance transcription of the locus by enhancing binding of BvgA (Parkhill, et al. 2003: 32-40). Thus, the *B. pertussis* promoter directs toxin expression (or greatly enhanced expression) and increased virulence. However, it is unclear whether these SNPs arose during speciation of *B. pertussis* or subsequent to this in a lineage that has likely displaced others, akin to the recent rise to dominance of *ptxP3*.

Whatever the scenario, it is very likely that *B. pertussis* is a recently evolved lineage, highly adapted (and possibly still adapting) to the human host, with little genetic diversity among the global population, certainly compared to many other bacterial pathogens. Fascinatingly, a very similar scenario has occurred twice more, giving rise to *B. parapertussis* (Parkhill, et al. 2003: 32-40) and *B. holmesii* (Harvill, et al. 2014). There is sparse information regarding these other two species. However, both have undergone IS element expansion in their genomes, both appear to have suffered genome reduction (likely a consequence of high numbers of IS repeats in the genomes) and both can cause Pertussis, or at least Pertussis-like disease. It is likely that comparative studies of these three *Bordetella* will yield a fresh insight into the evolution of whooping cough.

The availability of a reference *B. pertussis* genome sequence allowed further interrogation of *B. pertussis* diversity. Comparative genomic hybridisation using micro-arrays of numerous strains revealed a largely highly conserved gene content (Bouchez, et al. 2008: e2437, Kallonen, et al. 2011: 2034-42, King, et al. 2010: 64). Differences were largely due to the presence or absence of discrete regions, suggesting limited variability in IS element-derived deletions among strains. In general, recent strains contained more deletions than older strains suggesting that gene loss is on going in *B. pertussis*.

More recently, next generation sequencing has provided a quantum leap in analysis of *B. pertussis* enabling whole genome sequences to be generated for multiple strains, and for them to be resolved at single base resolution. Two recent studies in particular have highlighted the insights possible with this approach. A landmark study by Bart and colleagues generated and analysed genome sequences from 343 strains isolated over many decades from across the globe providing the definitive view of *B. pertussis* diversity and evolution over recent times (Bart, et al. 2013).

The limited genetic diversity of *B. pertussis* was confirmed with just 5 414 SNPs identified among all of the strains analysed. This allowed calculation of a mutation rate of 2.24×10^{-7} per site per year and an average SNP density of 0.0013 SNPs/bp. Interestingly, phylogenetic analysis suggested two distinct branches, Figure 1. Branch I contained just 6 strains and appears to represent a clade that is no longer observed, containing the seemingly older alleles ptxA5 and ptxP4. The two branches appear to have diversified approximately 2000 years ago representing either two independent introductions into the human population, or that lineages representing genotypes intermediate to the two branches have been lost. The genome reduction characteristic of *B. pertussis* occurred largely before diversification of the lineages as two reference strains, one from each lineage, shared 72% of pseudogenes with the inactivating mutations being identical between the two strains.

A potential role for vaccination in shaping the *B. pertussis* population is evident from these analyses. Most recent isolates are contained in a lineage, IIb, that arose prior to the introduction of vaccination but which has diversified since its introduction and this is associated with the rise of the *ptxA1* allele, Figure 1. Vaccination might be expected to reduce diversity as many lineages would be eradicated. However, if a particular genotype exhibits a degree of vaccine escape, it would expand and diversify in areas of vaccination. A second wave of diversification arose from the emergence and subsequent expansion of the lineage carrying *ptxP3*, discussed above. The introduction of ACV coincides with the decline of the *ptxA1* carrying lineage followed by further expansion of the *ptxP3* lineage by the acquisition of the *fim3-2* allele. These data do fit with the theory of vaccines promoting the expansion and subsequent diversification of particular genotypes, presumably those less controlled than others by vaccine-induced immunity. However, definitive evidence that these strains do resist vaccine-induced immunity to a greater degree than other lineages is currently lacking.

This seminal study included strains isolated during and subsequent to the introduction of ACVs but only 3 isolates were more recent than 2008. A more recent study built on these observations, again utilising whole genome sequencing but with a focus on strains isolated during the UK 2012 Pertussis outbreak (Sealey, et al. 2014). This revealed that this large outbreak was polyclonal in nature, with numerous closely related, but distinct clones contributing to it. Importantly, there was no novel genotype identified among the outbreak isolates. Indeed the isolates were very closely related to those observed among periods when the incidence of Pertussis was low. Thus, genomic data strongly suggests that resurgence is not the result of a particular genetic change producing a hypervirulent clone.

The genome sequences generated in these studies allowed calculation of SNP densities for different categories of genes, changes in these frequencies over time and correlation of changes with implementation of WCV or ACV. Bart et al observed that genes categorised as either ‘virulence’ or ‘transport and binding’ genes contained a SNP density higher than the genome average (Bart, et al. 2013). This could be due to these genes’ products being exposed to the bacteria’s environment and thus subject to e.g. immunity-mediated selection pressure. ACVs, containing just Prn, Ptx, Fim2, Fim3 and FHA, would be expected to induce selection pressure against just these five antigens. Sealey et al calculated that, compared to other genes encoding putative surface proteins, the genes encoding the ACV antigens had significantly higher SNP densities. Moreover, this was the case prior to the introduction of Pertussis vaccines but the SNP density in ACV antigen encoding genes is significantly higher among strains isolated since the introduction of ACVs compared to those isolated prior to this (Sealey, et al. 2014). This is consistent with the ACV antigen encoding genes being under selection pressure prior to vaccination. These proteins are highly immunogenic (hence being selected for vaccination) and infection would generate immunity directed against them. However, there is a clear signal that their rate of evolution has increased recently, and this could be due to ACV-induced immunity focusing on these few proteins. This raises obvious concerns about accelerated evolution of *B. pertussis* and the effect of this on the ability of ACV to control Pertussis.

These studies represent a step change in the analysis of *B. pertussis* and importantly establish genome sequencing as a central approach in this area. They have revealed unprecedented detail of *B. pertussis* diversity and evolution. Their findings support the hypotheses that vaccination creates selection pressure on *B. pertussis* and that certain genotypes have fitness advantages under this pressure. The consequence of this is presumably that *B. pertussis* is evolving towards ever increasing vaccine escape, with all of the public health concerns that this would entail. However, the data raise as many questions as they answer, and answering these questions is going to be critical to understanding the current picture of resurgence and perhaps more importantly, the likely trajectory of Pertussis incidence in future years.

For example, if vaccination creates selection pressure on ACV antigen genes, why is greater variation not observed? Although genomic analyses have identified changes in vaccine antigen gene alleles over time, and shown that these genes are faster evolving than other surface protein encoding genes, the level of variation appears to be still very low.

For *ptxA*, just eight alleles were observed world wide, including two silent mutations and thus just six putative protein variants are known, with just one or two amino acid differences between them. Furthermore, *ptxA1* and *ptxA2* were found in 96% of the strains (Bart, et al. 2013), including all of the recent UK isolates analysed (Sealey, et al. 2014). Five alleles, with four putative protein variants were observed for *fim3* (Bart, et al. 2013). Just two alleles are known for *fim2*, involving just a single amino acid change. FHA is not included in most analyses as this large gene (approximately 10.8kb) contains numerous repeat regions making mapping of short read data problematic. Interestingly, 14 alleles of the relatively short *ptxP* region were observed (Bart, et al. 2013) suggesting that sequence variation in *B. pertussis* does occur, although clearly the selection pressure must be different for this character compared to structural genes.

Compared to the very limited variation displayed by the other ACV antigens, the situation regarding Prn is different. In the global study by Bart et al, twelve alleles gave rise to 11 putative protein variants, although just 3 alleles were found in 92% of strains (Bart, et al. 2013). Unlike the other ACV antigens, variation in *prn* occurred mainly through variation in the number of repeats. This is perhaps a reversible process although the frequency at which repeat lengths might vary is unclear. Of particular importance is that very recently in several countries experiencing resurgence there has been a dramatic rise in the incidence of strains not expressing Prn, although with notable variation between countries. For example, in Finland, in a study of 76 clinical isolates, only 2 were pertactin-deficient by ELISA, both isolates came from infants who had received the acellular vaccine (Barkoff, et al. 2012: 1703-4). In Australia and Japan studies suggest that the proportion of pertactin negative strains in circulation is higher, with 27% of 121 strains tested in Japan (Otsuka, et al. 2012: e31985)and 30% of 320 strains tested in Australia negative for pertactin expression (Lam, et al. 2014: 626-33). In the US the prevalence of pertactin-deficient strains appears to be higher still, topping 50% of currently circulating strains (Pawloski, et al. 2014: 119-25). Prn deficiency arises by a number of mechanisms: gene deletions, SNPs creating internal stop codons and promoter inversion, but the most widespread mechanism is mutation by insertion of IS481 (Pawloski, et al. 2014: 119-25). Interestingly, Prn-deficiency was not observed among the global collection analysed by Bart et al (Bart, et al. 2013) of strains predominantly from 2008 or earlier, highlighting that the spread of these strains is a very recent occurrence. The rise in incidence of Prn-deficient strains, seemingly mainly in countries using ACVs, has been taken as a rare example of evidence for the selective pressure exerted by ACV-induced immunity. While there has been no reported difference in clinical presentation of disease between those infected with a strain of *B. pertussis* expressing pertactin and a pertactin-deficient strain, the likelihood of having reported disease caused by a pertactin-deficient isolate compared with one expressing pertactin is greater in people who have received the full course of pertussis vaccines, suggesting a selective advantage for pertactin-deficiency in areas where coverage by ACV is high (Martin, et al. 2015: 223-7). Furthermore, using a mouse model it was shown that pertactin deficiency confers a fitness advantage in an ACV-vaccinated host (Hegerle, et al. 2014: 6597-600). Naïve or ACV-vaccinated mice were infected with Prn-expressing or Prn-deficient strains. Naïve mice were colonized to the same levels regardless of Prn expression status, which is different to some previous reports (van Gent, et al. 2011:

e18014). However, Prn-expressing strains were cleared more rapidly from vaccinated mice compared to the Prn-deficient strains, implying that Prn-deficiency is a fitness advantage in ACV-vaccinated hosts (Hegerle, et al. 2014: 6597-600). It should be noted that the Prn-expressing and Prn-deficient strains very likely had other genetic differences, and thus to precisely identify the effect of Prn-deficiency, comparisons between WT and an isogenic Prn-deficient strain, or a naturally occurring Prn-deficient strain and one in which Prn expression has been reinstated, in ACV-vaccinated hosts should be made.

Thus, there appears to be striking differences between antigens in the effects of any ACV-mediated selection pressure. It is clear that Prn-deficient strains must be able to infect people and transmit between them, and as most isolates are from clinical cases they appear to be virulent in terms of causing disease. Thus, Prn must not be required for these processes. It is possible that the other antigens are required, preventing the establishment of strains deficient for any of them. However, it is harder to envisage that greater levels of polymorphism are not tolerated in either of Ptx, Fim2 or Fim3.

In particular, *B. pertussis* appears to be restricted to a single niche, the human nasopharynx. Many pathogens with similar restricted niches have much higher levels of variation within their populations than does *B. pertussis*. Even discounting the effect of vaccination, for continued circulation between hosts that have prior exposure to the pathogen there is likely selection pressure to vary in order to escape the immunity arising from previous exposure. Thus, other human restricted pathogens, for example *Neisseria meningitidis* and *Streptococcus pneumoniae* display extensive antigenic variation and exist as different clades with distinct gene repertoires. A similar scenario would appear to exist for *B. pertussis*. Does the small number of mutations observed in *B. pertussis* generate sufficient diversity for continuous circulation among its human hosts? Testing this is a very difficult task. Generating strains carrying individual SNPs to enable the effect of specific mutations to be analysed is feasible. However, such studies have been hindered by the lack of suitable models of disease and transmission. The infant baboon model does enable study of these aspects but experiments with these hosts are limited to low numbers of animals and subtle effects may be missed due to this. In vitro assays, for example measuring the effect of individual SNPs on recognition of strains by antibodies in vaccine-immune sera or stimulation by of T cells from vaccinated individuals, have the capacity to be high throughput enabling the testing of many combinations of SNPs and sera. However, although both antibodies and T cells are likely to be important for protective immunity, definitive correlates of protection have yet to be identified making it difficult to extrapolate results from in vitro assays to the in vivo situation.

Alternatively, is it possible that host immunity is not a strong selective pressure on *B. pertussis*? A number of recent studies have indicated that both natural infection induced- and vaccine-induced immunity is somewhat short-lived, waning by adolescence, if not earlier. For vaccine-mediated immunity in particular to be a strong selective pressure on *B. pertussis*, pre-adolescent hosts must be a major part of the ecology of *B. pertussis*. Little is known about the circulation of *B. pertussis* in the general population. Very few studies have monitored *B. pertussis* outside of people presenting with cough symptoms. It is reasoned that anti-PT antibody titres above a certain threshold in people with no recent history of cough illness or vaccination are an indication of recent infection by/exposure to *B. pertussis*. Seroprevalence levels vary widely between studies. However, many suggest that many more people encounter *B. pertussis* than present with disease, for example (Huygen, et al. 2014: 724-8, Ronn, et al. 2014: 729-37, Scott, et al. 2015: 333-8). In this case it is likely that *B. pertussis* is endemic in the human population but causes disease in only a proportion of those carrying the organism, with infants being particularly susceptible to disease in its most serious form. If asymptomatic infection in 'healthy' people is the major *B. pertussis*-human interaction then it is critical to understand the nature of that interaction, particularly the host response, to determine possible immune-mediated selection pressures that might be exerted on the bacteria. In this regard, it is also important to isolate bacteria from cases of asymptomatic carriage. To date, all isolates examined have been recovered from cases of Pertussis, or at least Pertussis-like cough illness. It is important to determine that these isolates are representative of the larger *B. pertussis* population and that disease-causing *B. pertussis* are not some separate lineage, although there is no evidence that this is the case.

An additional scenario is that *B. pertussis* does exhibit variation, but this is not via variation in DNA content or DNA sequence *per se*. Genomic analyses revealed variability among strains in the deletions that they carried (Bouchez, et al. 2008: e2437, Kallonen, et al. 2011: 2034-42, King, et al. 2010: 64). While a number of deletions appear to be common to many strains, numerous strains carry either strain-specific deletions, or deletions shared by just a few strains (for example, (Sealey, et al. 2014)). It has been assumed that deleted regions contain genes not required for *B. pertussis's* human-adapted niche, and thus that deletions are mainly neutral in terms of fitness, but this has not been investigated. In addition, the first genome sequence of *B. pertussis* Tohama I revealed the role of IS481-mediated recombination not only in generating deletions but also in intra-genomic recombination (Parkhill, et al. 2003: 32-40). Little is known about conservation or variability of IS481 insertions among strains, and thus the potential for variation in recombination and the resulting genome arrangement among them. Although genome sequences have been generated for hundreds of strains, this was largely using Illumina, short read sequencing. These sequence reads are too short to span the 1kb IS481 repeats and thus genome assembly programmes are confounded by the repeats, producing assemblies comprised of over 200 contigs, most of which likely terminate at an IS481 element. Thus, although the DNA content of each strain and the sequence of this content is identified by this approach, it is not possible to produce closed, ordered genome sequences from these data. The advent of long-read sequence platforms, including the PacBio RS system and potentially the Oxford Nanopore technology, enables the generation of reads that encompass IS repeats and allow long range scaffolds to be assembled and closed genome sequences to be achieved (Liao, et al. 2015: 8747). The closed genome sequences of two Dutch strains assembled in this way identified three large inversions when comparing the two sequences, demonstrating different genome arrangements between the two strains (Bart, et al. 2014). Comparison of additional strains for which closed genome sequences have been generated identifies a greater level of genome arrangement variation Figure 2 (Preston unpublished). In this alignment based analysis, numerous inversions and translocations are apparent. That such a range of genome arrangements was found among the limited number of strains analysed suggests that the potential for genome arrangement variation among *B. pertussis* strains could be extensive.

Genome arrangement affects several factors thought to be determinants of gene expression (for example see (Couturier and Rocha 2006: 1506-18): distance of a gene from the origin/terminus of DNA replication, the strand on which the gene is encoded (leading or lagging), transcriptional organisation (e.g. operon structure, orientation to neighbours), the distance of transcriptional regulators from their target genes, the organisation of non-coding RNAs and the location of genes in relation to transcriptionally active IS elements. Even genetically identical cells can exhibit phenotypic variations arising from stochasticity in gene expression. Extrinsic noise, arising from e.g. fluctuations in cellular components can give rise to fluctuations in gene expression (Elowitz, et al. 2002: 1183-6, Raser and O'Shea 2005: 2010-3). Variation in genome arrangement is likely to generate different internal environments between strains, affect gene expression and lead to variation in phenotype. Thus, it is possible that variation in genome arrangement is generates diversity among *B. pertussis*, but this remains speculation at present.

The difficulty will be devising approaches to investigate diversity among strains. Many of the detected variations are within uncharacterised genome regions, meaning it is not clear what phenotypic traits might vary, and that it is unlikely the necessary assays exist. A major obstacle is that subtle variations likely lead to subtle variations in phenotypes. Studies in traditional models such as murine or baboons can accommodate only small numbers of hosts meaning that the power to detect minor differences is lacking, although at the population level (of hosts) subtle variations may be significant. If variation in genome arrangement does affect gene expression levels then they are detectable using gene expression profiling (e.g. RNAseq, see below) but the difficulty then becomes interpretation of these complex datasets and still leads to the same problem of testing hypotheses raised by such studies regarding phenotypic variation.

Outlook

The resurgence of Pertussis has led to renewed interest in *B. pertussis*. It is evident that our understanding of the infection biology of this bacterium, at both the level of an individual host, and at

the population level is surprisingly poor. The advent of vaccination appeared to remove the need to study *B. pertussis* as a pressing infectious disease issue and while very elegant work has focused on the biology of the major 'virulence factors', it is clear that our knowledge does not explain sufficiently what is happening at present.

Clearly it is very important to decipher the precise mechanisms of immunity arising from infection and vaccination (both WCV and ACV). This will very likely shed light on the role of the switch to ACV in resurgence. This work will also define precise immune correlates of protection which will greatly aid developments involving changes to vaccine schedules or the development of enhanced or new vaccines. It is also important to understand whether changes to *B. pertussis* observed over the last decades is adaptation to host factors, particularly vaccine-induced immunity, and to be able to extrapolate this to the future to enable policy decisions to be made regarding future control measures for Pertussis. In turn this will involve also a much deeper understanding of the epidemiology of Pertussis, as this is required to target control measures to the right members of the population to achieve the best overall protection possible. Genomics-based approaches will play a key role in all of these areas.

Monitoring further changes in *B. pertussis* is clearly important for identifying and understanding genetic changes that might herald further changes in behaviour and further adaptations, whether in response to vaccination or otherwise. It is clear that fine scale resolution between strains, and capture of all of the genetic variation among *B. pertussis*, requires whole genome sequencing approaches. The high throughput capacity, decreasing cost and increasing accessibility of next generation DNA sequencers to researchers and reference labs means that WGS approaches will likely become routine for surveillance of *B. pertussis*, an approach that will still enable traditional characters such as MVLA type and allele types to be ascertained. However, it is worth considering that by far and away the most common NGS platform used is Illumina, based on short read sequencing. If ordered, assembled genome sequences are required to fully appreciate *B. pertussis* diversity then longer read technologies will need to be considered incorporated into this surveillance. The issue then arises of how to utilise such data into a useable epidemiological framework. Would isolates that differ by just a single SNP, or those that have no SNP differences but differ by a translocation be regarded as different strains, or what level of variation would be considered a different lineage? If the aim of this surveillance is to identify changes that impact on infection then the same problem arises as discussed above, i.e. the difficulty in deciphering genetic changes in terms of assayable phenotypes, and the lack of models that might detect subtle changes. A further challenge to continued surveillance of *B. pertussis* is the increasing use of non-culture based diagnosis, in particular PCR and serology. These approaches have the great advantage of being relatively non-invasive, and fast, compared to the nasopharyngeal swabs required for culture-based diagnosis. However, it is crucial that collection of *B. pertussis* cultures is continued in order to be able to monitor the pathogen, detect further changes in it and for studies that investigate the effects of any changes.

The resurgence of Pertussis has given renewed impetus to the need to better understand *B. pertussis* infection biology, including the host response. Despite the deciphering of the genetic code of *B. pertussis* over 10 years ago, the number of functionally characterised loci in *B. pertussis* remains remarkably small. These loci include those coding for adhesins, toxins, and other factors important for infection but these represent a very small fraction of the gene content of *B. pertussis*. Functional genomics offers a means to move away from traditional reductionist approaches of studying individual gene systems at a time towards systems level analyses of *B. pertussis* during infection. Transcriptional profiling is a powerful approach for genome-wide monitoring of active processes, and changes in gene expression levels are a valuable indicator of those processes whose activity varies e.g. during the course of an infection, or dependent on the within-host site occupied. While transcriptional profiling is admittedly just a proxy for protein expression, the relatively tight correlation in bacteria between transcriptional and translational levels makes it a very useful readout. Thus, transcriptional profiling of bacteria in different host niches within the respiratory tract, over time, during infection will shed new light on the activity of characterised systems, and identify novel systems that are tightly linked to the progression of Pertussis from initial colonisation to symptomatic disease and transmission to new hosts. The infant baboon model is likely to

be central to these advances. Applying this approach to the study of host responses to infection will be valuable to defining the responses of host cells to infection and immunisation and in turn for deciphering the mechanisms operating for each. This will be critical to not only understanding the current situation, but also in the design of potential novel *B. pertussis* vaccines, the need for which has been hotly debated (Poolman 2014: 1159-62).

The resurgence of Pertussis in countries using the ACV has raised serious public health concerns. It is not yet clear whether resurgence is temporary or if recent trends signal a sustained period of (relatively) high levels of disease. There is strong evidence implicating the use of ACVs as a contributing factor to resurgence. A number of solutions have been discussed including various programmes of additional boosters using ACVs, the inclusion of additional antigens or adjuvants to the ACV as well as possible development of novel Pertussis vaccines. In addition, evidence for *B. pertussis* evolution under the influence of vaccine-mediated immunity means that the pathogen side of the equation must be factored in to these scenarios. From this, it is clear that our current understanding of *B. pertussis* infection and immunity is insufficient to allow a solution to be devised with any certainty of its success. Remediation of this knowledge gap will be heavily dependent on the use of genomics-based approaches.

References.

Barkoff AM, Mertsola J, Guillot S *et al.* Appearance of Bordetella pertussis strains not expressing the vaccine antigen pertactin in Finland. *Clinical and vaccine immunology : CVI* 2012;**19**: 1703-4.

Bart M, Harris S, Advani A. Global Population Structure and Evolution of Bordetella Pertussis and their Relationship with Vaccination. *submitted* 2013.

Bart MJ, Zeddeman A, van der Heide HG *et al.* Complete Genome Sequences of Bordetella pertussis Isolates B1917 and B1920, Representing Two Predominant Global Lineages. *Genome announcements* 2014;**2**.

Bouchez V, Caro V, Levillain E *et al.* Genomic content of Bordetella pertussis clinical isolates circulating in areas of intensive children vaccination. *PloS one* 2008;**3**: e2437.

Burns DL, Meade BD, Messionnier NE. Pertussis Resurgence: Perspectives From the Working Group Meeting on Pertussis on the Causes, Possible Paths Forward, and Gaps in Our Knowledge. *Journal of Infectious Diseases* 2014;**209**: S32-S5.

Cherry JD. Pertussis - the Trials and Tribulations Of Old and New Pertussis Vaccines. *Vaccine* 1992;**10**: 1033-8.

Cherry JD. Acellular pertussis vaccines-a solution to the pertussis problem. *The Journal of infectious diseases* 1993;**168**: 21-4.

Cherry JD. Historical Review Of Pertussis and the Classical Vaccine. *Journal Of Infectious Diseases* 1996;**174**: S 259-S 63.

Couturier E, Rocha EP. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Molecular microbiology* 2006;**59**: 1506-18.

Darling AC, Mau B, Blattner FR *et al.* Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome research* 2004;**14**: 1394-403.

Diavatopoulos DA, Cummings CA, Schouls LM *et al.* Bordetella pertussis, the Causative Agent of Whooping Cough, Evolved from a Distinct, Human-Associated Lineage of B. bronchiseptica. *PLoS pathogens* 2005;**1**: e45.

Elowitz MB, Levine AJ, Siggia ED *et al.* Stochastic gene expression in a single cell. *Science* 2002;**297**: 1183-6.

Harvill ET, Goodfield LL, Ivanov Y *et al.* Genome Sequences of Nine Bordetella holmesii Strains Isolated in the United States. *Genome announcements* 2014;**2**.

Hegerle N, Dore G, Guiso N. Pertactin deficient Bordetella pertussis present a better fitness in mice immunized with an acellular pertussis vaccine. *Vaccine* 2014;**32**: 6597-600.

Huygen K, Rodeghiero C, Govaerts D *et al.* Bordetella pertussis seroprevalence in Belgian adults aged 20-39 years, 2012. *Epidemiology and infection* 2014;**142**: 724-8.

Kallonen T, Grondahl-Yli-Hannuksela K, Elomaa A *et al.* Differences in the genomic content of Bordetella pertussis isolates before and after introduction of pertussis vaccines in four

European countries. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2011;**11**: 2034-42.
King AJ, van der Lee S, Mohangoo A *et al.* Genome-wide gene expression analysis of Bordetella pertussis isolates associated with a resurgence in pertussis: elucidation of factors involved in the increased fitness of epidemic strains. *PloS one* 2013;**8**: e66150.
King AJ, van Gorkom T, van der Heide HG *et al.* Changes in the genomic content of circulating Bordetella pertussis strains isolated from the Netherlands, Sweden, Japan and Australia: adaptive evolution or drift? *BMC genomics* 2010;**11**: 64.
Kurniawan J, Maharjan RP, Chan WF *et al.* Bordetella pertussis clones identified by multilocus variable-number tandem-repeat analysis. *Emerging infectious diseases* 2010;**16**: 297-300.
Lam C, Octavia S, Bahrame Z *et al.* Selection and emergence of pertussis toxin promoter ptxP3 allele in the evolution of Bordetella pertussis. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2012;**12**: 492-5.
Lam C, Octavia S, Ricafort L *et al.* Rapid increase in pertactin-deficient Bordetella pertussis isolates, Australia. *Emerging infectious diseases* 2014;**20**: 626-33.
Lapin JH. *Whooping cough*. Springfield, Ill: Charles C. Thomas, 1943.
Liao YC, Lin SH, Lin HH. Completing bacterial genome assemblies: strategy and performance comparisons. *Scientific reports* 2015;**5**: 8747.
Litt DJ, Neal SE, Fry NK. Changes in genetic diversity of the Bordetella pertussis population in the United Kingdom between 1920 and 2006 reflect vaccination coverage and emergence of a single dominant clonal type. *Journal of clinical microbiology* 2009;**47**: 680-8.
Lugauer S, Heininger U, Cherry JD *et al.* Long-term clinical effectiveness of an acellular pertussis component vaccine and a whole cell pertussis component vaccine. *Eur J Pediatr* 2002;**161**: 142-6.
Martin SW, Pawloski L, Williams M *et al.* Pertactin-negative Bordetella pertussis strains: evidence for a possible selective advantage. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2015;**60**: 223-7.
Mira A, Pushker R, Rodriguez-Valera F. The Neolithic revolution of bacterial genomes. *Trends in microbiology* 2006;**14**: 200-6.
Mooi FR. Bordetella pertussis and vaccination: the persistence of a genetically monomorphic pathogen. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 2010;**10**: 36-49.
Mooi FR, Van Der Maas NA, De Melker HE. Pertussis resurgence: waning immunity and pathogen adaptation - two sides of the same coin. *Epidemiology and infection* 2014;**142**: 685-94.
Mooi FR, van Loo IH, van Gent M *et al.* Bordetella pertussis strains with increased toxin production associated with pertussis resurgence. *Emerging infectious diseases* 2009;**15**: 1206-13.
Otsuka N, Han HJ, Toyoizumi-Ajisaka H *et al.* Prevalence and genetic characterization of pertactin-deficient Bordetella pertussis in Japan. *PloS one* 2012;**7**: e31985.
Parkhill J, Sebaihia M, Preston A *et al.* Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. *Nat Genet* 2003;**35**: 32-40.
Pawloski LC, Queenan AM, Cassiday PK *et al.* Prevalence and molecular characterization of pertactin-deficient Bordetella pertussis in the United States. *Clinical and vaccine immunology : CVI* 2014;**21**: 119-25.
Poolman JT. Shortcomings of pertussis vaccines: why we need a third generation vaccine. *Expert review of vaccines* 2014;**13**: 1159-62.
Preston A, Parkhill J, Maskell DJ. The Bordetellae: lessons from genomics. *Nat Rev Microbiol* 2004;**2**: 379-90.
Raser JM, O'Shea EK. Noise in gene expression: origins, consequences, and control. *Science* 2005;**309**: 2010-3.
Rendi-Wagner P, Kundi M, Mikolasek A *et al.* Hospital-based active surveillance of childhood pertussis in Austria from 1996 to 2003: estimates of incidence and vaccine effectiveness of whole-cell and acellular vaccine. *Vaccine* 2006;**24**: 5960-5.
Ronn PF, Dalby T, Simonsen J *et al.* Seroepidemiology of pertussis in a cross-sectional study of an adult general population in Denmark. *Epidemiology and infection* 2014;**142**: 729-37.

Ross PJ, Sutton CE, Higgins S *et al.* Relative contribution of Th1 and Th17 cells in adaptive immunity to Bordetella pertussis: towards the rational design of an improved acellular pertussis vaccine. *PLoS pathogens* 2013;**9**: e1003264.
Scott S, van der Sande M, Faye-Joof T *et al.* Seroprevalence of pertussis in the Gambia: evidence for continued circulation of bordetella pertussis despite high vaccination rates. *The Pediatric infectious disease journal* 2015;**34**: 333-8.
Sealey KL, Harris SR, Fry NK *et al.* Genomic Analysis of Isolates From the United Kingdom 2012 Pertussis Outbreak Reveals That Vaccine Antigen Genes Are Unusually Fast Evolving. *The Journal of infectious diseases* 2014, DOI 10.1093/infdis/jiu665.
van Gent M, Bart MJ, van der Heide HG *et al.* Small mutations in Bordetella pertussis are associated with selective sweeps. *PloS one* 2012;**7**: e46407.
van Gent M, Heuvelman CJ, van der Heide HG *et al.* Analysis of Bordetella pertussis clinical isolates circulating in European countries during the period 1998-2012. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2015;**34**: 821-30.
van Gent M, van Loo IH, Heuvelman KJ *et al.* Studies on Prn variation in the mouse model and comparison with epidemiological data. *PloS one* 2011;**6**: e18014.
Warfel JM, Beren J, Kelly VK *et al.* Nonhuman primate model of pertussis. *Infection and immunity* 2012;**80**: 1530-6.
Warfel JM, Edwards KM. Pertussis vaccines and the challenge of inducing durable immunity. *Curr Opin Immunol* 2015;**35**: 48-54.
Warfel JM, Zimmerman LI, Merkel TJ. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. *Proceedings of the National Academy of Sciences of the United States of America* 2013, DOI 10.1073/pnas.1314688110.
WHO. Global and regional immunization profile volume 2015, 2014.
Witt MA, Katz PH, Witt DJ. Unexpectedly limited durability of immunity following acellular pertussis vaccination in preadolescents in a North American outbreak. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2012;**54**: 1730-5.

Figure Legends.

Figure 1.
Phylogeny of a global collection of *B. pertussis* strains depicting the two lineages denoted as Branch I and II. The approximate periods in which WCV were used widely and in which ACV were introduced are indicated. The appearance of prominent alleles is indicated on the tree and the ACV antigen allele profile of strains is colour-coded. Green circles indicate the two reference strans Tohama I (T) and 18323; black circles indicate two US vaccine strains (B308, branch I and B310, branch II). Reproduced from (Bart, et al. 2013) with permission.

Figure 2.
Variation in genome arrangement among *B. pertussis* strains. Alignment of closed, ordered genome sequences of multiple *B. pertussis* strains was performed using Mauve (Darling, et al. 2004: 1394-403). This displays aligned genomes as locally colinear blocks (LCBs); conserved segments of genomes that are free from internal rearrangements. The *B. pertussis* Tohama I genome sequence (Parkhill, et al. 2003: 32-40) is displayed as the reference sequence. In other strains, an LCB that is located at a different point in the genome indicates a translocation relative to Tohama I. An LCB below the central black line indicates an inversion event relative to Tohama I. The alignment also displays relative DNA sequence conservation where major deletions are shown as white sections within LCBs. To note here are that each strain contains rearrangements relative to Tohama I and each other. Strains B1920 and B1917 are Dutch strains isolated in 2000 representing the major *ptxP1* and *ptxP3* lineages respectively (Bart, et al. 2014). UK strains have been described previously (Sealey, et al. 2014) and were isolated in 1920 (UK1), 1942 (UK2), 2006 (UK16), 2011 (UK51) and 2012 (UK69). The closed genome sequences were generated using both PacBio and Illumina sequencing (Preston, unpub.). D420 is a US isolate (Warfel, et al. 2012: 1530-6).